Using Physical Context for Just-in-Time Information Retrieval

Bradley Rhodes

Abstract—Jimminy is a wearable personal note-taking and note-archival application that automatically displays notes that might be relevant to the wearer in his current environment. The system selects old notes to show on a head-up display based on the wearer's current location, people in the immediate area, and the subject-line and contents of any current notes being written. This paper describes an experiment that evaluates the usefulness of the wearer's physical context (location and people in the area) for automatically finding useful archived information. The results suggest that, while physical context can be used to discover useful archived notes, the subject and text of notes currently being entered are a much better indicator of usefulness in the personal note-taking domain.

Index Terms—Wearable computers, intelligent agents, information retrieval, context.

1 Introduction

A Just-in-time Information Retrieval (JITIR) agent is software that proactively retrieves and presents information based on a person's local context. JITIRs can be thought of as automatic "query-free" search engines. Rather than using a human supplied query, the JITIR uses elements of a person's environment (physical or computational) as an automatically executed query into a database of potentially useful information. Any information that passes a relevance threshold is automatically presented to the user, usually via an interface that minimizes distraction should the information not be relevant after all.

There are several desktop-based applications that meet the definition of a JITIR, ranging from domain-specific to general information applications. One domain-specific application is FixIt [1], a system that delivers repair-manual entries to a technician working with diagnostic software for copier repair. Another is Coach [2], which provides automatic help with LISP programming within the context of an integrated development environment. Lumiere [3], the basis for the Microsoft Office Assistant, is an example of a more general JITIR that provides help using a large number of Microsoft products based on Bayesian user models. More general still are Letizia [4], which suggests links to follow based on a user's Web-browsing activities, Watson [5], which automatically performs Web searches based on text being written or read in Microsoft Word, the Remembrance Agent [6], which suggests personal email and documents based on text being written or read in Emacs, and SUITOR [7], which displays stock market quotes, headlines, and other information based on text in a word processor or Web browser.

All the applications described above select information to display based on a user's computational environment: text being written or read or interactions with a particular application. Wearable and mobile computers provide the opportunity to perform the same kind of query-free retrieval of information using the wearer's physical context such as his location, people in the area, or date instead of only using the user's computational environment. However, most of the wearable and mobile applications of this kind are quite domain-specific and focus on providing information that is especially suitable for indexing by the physical features used in the system.

Location is an especially popular physical feature for automatic presentation of information on a wearable system. Examples include the Touring Machine [8] and Cyberguide [9], which automatically provide information about nearby places of interest in a city using GPS data. Similarly, the CIS Archeologist Field Assistant [10] uses GPS data to automatically retrieve field data collected about giraffes within a Kenyan game reserve. On a smaller scale, several wearable and hand-held systems use location to deliver information about exhibits at a conference or museum [11], [12], [13].

Systems using physical context other than location have also been developed. The DyPERS system [14] also presents information about museum exhibits, but, instead of location, uses machine vision to detect what painting a wearer of the system is currently viewing. The Augment-able Reality system uses two-dimensional bar codes attached to objects to bring up information whenever that object is in view [15]. Finally, systems are being developed to display biographical information about a person based on face-recognition machine-vision software [16].

The mobile and wearable systems described above all use a particular physical context to automatically retrieve information intimately associated with that context. It is not clear, however, how the use of physical context as an automatic query can generalize to broader domains. In particular, it is not clear whether physical context such as a person's location or people in the area is useful for automatically retrieving personal notes in an office setting or from a person's daily life.

The Jimminy system, also known as the Wearable Remembrance Agent, was designed to test the utility of physical context cues for automatically retrieving information in a personal note-taking system. Unlike the mobile systems described above, Jimminy displays information that is not explicitly chosen to fit well with the physical context being used for retrieval. The question being addressed is how well physical context can work when the association between information and physical context in which the information was obtained is only implicit, as in the case of personal notes in an office or home environment.

1.1 Jimminy (the Wearable Remembrance Agent)

Jimminy is a wearable system that senses a wearer's location, people in the area, and text being entered on a wearable computer, and automatically displays information that might be relevant in that environment on a head-up display. The program has two concurrently running components, the note-taking environment and the automatic retrieval environment. Both components operate within the Emacs text-editor running on a Lizzy wearable computer [17], outfitted with a Private Eye head-up display, a Twiddler one-handed chording keyboard, a radio-frequency receiver, and an infrared-receiver. Jimminy is based on a desktop version called the Remembrance Agent [18], [6], which has the same basic functionality but does not use any physical sensors. An earlier version of Jimminy is described in [19].

Notes are entered using the one-handed keyboard and can be touch-typed at a rate between 35 and 50 words per minute. New notes are automatically tagged with a header indicating the wearer's current location, people in the area, and time and date. A subject line can be manually entered. Notes are displayed in the top 20 lines of the 320×240 monochrome head-up display, using 80 percent of the screen real-estate.

The automatic retrieval component of Jimminy continuously watches the environment (wearer's location, people in the immediate vicinity, subject of any notes being written, and the text of any notes being written) and automatically displays up to five one-line summaries of past notes that might be relevant to the

- *The author is with Ricoh Innovations, 2882 Sand Hill Rd., Menlo Park, CA 94025-7054. E-mail: rhodes@bradleyrhodes.com.*

Manuscript received 1 July 2002; revised 1 Dec. 2002; accepted 1 Feb. 2003. For information on obtaining reprints of this article, please send e-mail to: tc@computer.org, and reference IEEECS Log Number 118309.

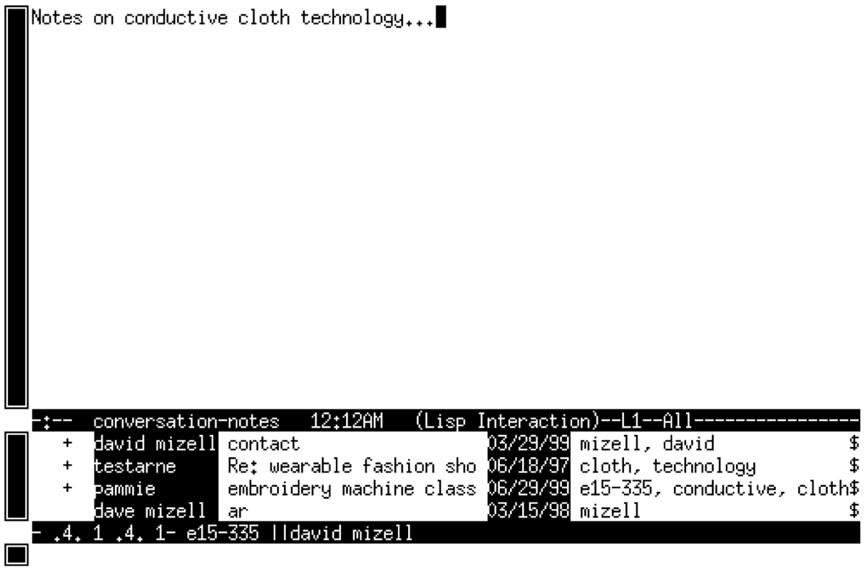

Fig. 1. The Jimminy agent screenshot.

current situation, given the sensed context (see Fig. 1). When the environment is changing (for example, when the wearer moves to a new location or takes notes), the list of suggestions is updated every five seconds. The head-up display is positioned such that suggestions can be scanned easily but are not distracting. If desired, the full text of a suggested note can be retrieved with a single chord on the one-handed keyboard.

Notes are suggested based on similarity to the current environment. For example, upon entering a room, notes that were previously taken in that room are suggested. When a new person enters the room, notes that were previously taken with that person present will also be shown, with preference to notes where both the location and person match. As a note is typed, the text of the note is compared to the text of previous notes using a Term Frequency/inverse Document Frequency (TF/iDF) algorithm that retrieves notes on a similar topic [20]. The four different dimensions (location, people, subject, and text) are all used to compute relevance and the four similarity results are linearly combined to form an overall ranking of potentially relevant documents. Earlier versions of Jimminy also used date, time, and day-of-week as context information, though more recent versions do not. Dimensions that have recently changed (for example, "location" when a new room has just been entered) receive extra weight when computing similarity.

Jimminy detects the wearer's location and people in the area using radio location beacons and infrared active name badges. The active name badges use the Locust IR beacon [21], which has a directional range of a few feet. Location beacons are Locusts that have been modified to transmit radio instead of infrared and are omnidirectional with a radius of about 10-15 feet. At the time of the experiment, location beacons were placed in several rooms around the MIT Media Lab, with one or two per room. Whenever the wearer of the system comes into range of a beacon, Jimminy automatically updates its record of the user's location and who is in the area. Sensors are integrated using a Java-based distributed agent architecture called Hive [22], [23], which handles the low-level beacon protocol and translation from beacon IDs into location and person names.

Jimminy was in use at the MIT Media lab for several years, but beacons and active badges were not always available and were not the primary focus of the research. Location beacons were not distributed beyond a few rooms within the MIT Media Lab and never outside the lab. Active name badges were not used except for demonstrations. To handle cases when location and person beacons were not available, physical information could be entered manually using the one-handed chording keyboard. The keyboard was also used to enter notes and to give each note a title or subject line. Time and date-stamps were automatically added to a note as it was created. Due to the lack of beacon coverage, manual entry was by far the most common way Jimminy received knowledge about the wearer's physical context.

2 Evaluation

Jimminy is designed for the recording and suggesting of personal notes, and collections of such notes take time to build and will be greatly affected by the environment in which they are taken. To understand how the system might be used over long periods of time in real-world environments, Jimminy was worn and used by the primary researcher in his daily life over the course of several years. This was a part of the MIT Media Lab living experiment in wearable computing.

From 1996 to 2000, Jimminy was used daily by the author and more than 850 notes were written and annotated on topics ranging from classes and conversations at conferences to descriptions of dance steps. Of these, 664 were tagged with information about the physical context in which the note was taken. The remaining notes were plain-text files without additional tags.

The quantitative evaluation of Jimminy was designed to test the value of suggestions and, in particular, to test whether information about location and people in the area improved suggestions.

2.1 Method

Six different sets of paired notes were generated using the Jimminy automatic-retrieval system: Location, Person, Subject, Note-text, All-features, and Control group. Each set was generated by processing notes that have already been written through the Jimminy automatic-retrieval system as if that note was just being entered. The first note of each pair was designated the "query" note, the note suggested by Jimminy was the result note. Before generating the suggestion pairs, the weights for Jimminy were adjusted so only the desired pieces of context would be used in generating a suggestion. For example, pairs in the Location set were generated only by looking only at similarity in the location field of a note's header. Similarly, the Note-text set was generated by taking every note and pairing it with the top suggestion

TABLE 1
Scores for Each Set

	Number scores tallied					
Set	1	2	3	4	5	Percentage scores 4 or 5
All Features	16	4	2	11	17	56%
Note Text	16	2	7	11	14	50%
Subject	32	5	1	5	7	24%
Location	37	6	1	2	4	12%
Person	40	3	3	2	2	8%
Control	45	3	2	0	0	0%

generated by Jimminy when all physical context and the subject was ignored. The All-features set was generated by pairing each note with whatever note would normally be suggested by Jimminy, using all the physical context as well as the subject and text of the note. The Control group consisted of random pairings. For each set, notes that were blank in the given field were removed. For example, the Person set had only notes that were tagged with at least one person.

A test suite was then created by taking 50 pairs at random from each of the six sets, for a total of 300 pairs. Pairs were then presented in random order and without labels to the researcher who originally took the notes. The researcher evaluated each pair for usefulness based on the following question: "If one of these notes were being written right now, how likely would it be for the other note to be useful?" Notes were rated one through five, with one being "definitely useless," four being "probably useful," and five being "definitely useful."

The primary advantage to the methodology described above is that evaluations are performed using both "query" environments and a notes database that actually occurred during long-term use. Furthermore, the fact that notes were taken over the course of several years makes it possible to evaluate long-term effects such as returning to the same classroom for a new class during a different school-year. However, the need for long-term usage also makes it difficult to run tests with multiple subjects. The results described here are only drawn from a single subject: the researcher himself while he was a 26-30-year-old graduate student. Results would likely differ for notes taken by subjects with other backgrounds, such as traveling salesmen or law enforcement authorities.

It should also be noted that the real usage of Jimminy differs from the test described above in a few important ways. First, Jimminy shows not one but five suggestions at any one time, but this experiment only rated the topmost suggestion that would be displayed. Second, if a suggestion's similarity rating is below a certain threshold, then Jimminy does not display it at all. This experiment scored all top suggestions, even those that would not have been displayed due to poor confidence in the suggestion. Also, cases where no suggestion could be found were automatically given the lowest score. Third, suggestions that were computed using the body of a note (the All-features set and the Note-text set) were computed using the entire text of the note. In actual use, these suggestions would only appear after the user had typed in the entire note; suggestions that would be displayed when the note was only halfway entered may be different. Finally, this experiment was conducted over the entire notes collection, which means notes might be suggested that were actually taken after the "query" note had been written. In an actual system, the first notes written about a new topic will necessarily produce no related notes. Only after other notes are taken on the same topic will useful suggestions arise.

TABLE 2
Breakdown of Location and Person Features

	Location	Person
Notes with feature not blank	558	466
Notes with unique feature (singleton)	89	167
Notes with most common	74	33
Notes with 2nd most common	65	27
Notes with 3rd most common	55	11
% Notes with unique feature	16%	36%
% Notes with most common	13%	7%
% Notes with one of two most common	25%	13%
% Notes with one of three most common	35%	15%

2.2 Results and Discussion

Table 1 lists scores tallied for each of the six groups and the percentage of pairs that were rated either "probably useful" or "definitely useful." The difference between the five experimental groups and the control group are all statistically significant ($p = 0.05$). The difference between the All-features set and the Note-text set was not significant.

Several conclusions can be drawn from this experiment. First, it is clear that the standard algorithm for Jimminy (the All-features set) produces useful suggestions, at least for this particular user and this particular set of notes. This is a promising result and supports the user's subjective opinion that the system produced useful suggestions during actual use.

Second, it is clear that the location in which a note was taken and people who were present at the time are not nearly as useful as the contents of a note or the subject line for determining what information might be useful to display. While a hit rate of around 10 percent may be good enough for some applications, it is still disappointing.

Third, the use of location, person, subject, and note text did not significantly improve the relevance of suggestions beyond that obtained using only the note text. This lack of improvement was not entirely due to the physical criteria having no effect on suggestions: 40 percent of the queries gave different results in the All-features condition than they did in the Note-text condition.

One likely cause for the poor performance is that the location and people in the area are poor distinguishing features for this particular set of notes. To be a good feature for distinguishing between notes, a feature should neither occur only once (such that no suggestions are produced) nor should it occur so frequently that the system is left with too many notes to choose from after filtering based on the feature. As a graduate student, the researcher spent very little time outside of his office or the two classrooms within the Media Lab. He also tended to meet a large number of people at conferences, demos, or talks that were never seen again.

Table 2 shows the breakdown of how frequently locations and people appeared in the corpus of notes. Locations were especially clumped, with over a third of the notes being taken in one of three places. These locations were, in order of frequency, the common area just outside the note-taker's office, his office itself, and the main classroom for the Media Lab. This is similar to problems reported with using location as a cue for query-based retrieval in the PEPYS system [24], where many unrelated activities occurred when the system user was alone in his office [25]. The people feature had the reverse problem: Over a third of the notes taken were tagged with a person that never appeared again in any other notes. This is likely due to frequent sponsor demos and guest lecturers that occur at the Media Lab. In both these situations, a person was often met once and never seen again.

While the distribution of locations and people within the notes database clearly contributes to the poor showing for these features, there is another explanation: The location where a note was taken and people present are simply a poor indicator of the topic of notes taken for this experiment. The assumption is that notes on the same topic as the wearer's current conversation, lecture, meeting, or idea will be useful. The problem is that "topic" is hard to define and even harder for a machine to sense and represent. The features that Jimminy actually uses (location, person, subject, and note text) are all used as proxies for the topic of the wearer's current thought. It stands to reason that the best representation of this topic is the text of the notes he is taking at the time, especially since this feature is the richest representation of the four. It also stands to reason that the subject line of a note will be a reasonably good proxy for topic since the subject line is composed by the note-taker for that purpose. The person and location fields do not have the richness of the note text nor are they chosen specifically by the note-taker to be good representations of topic. This does not mean the physical context cannot still be a good proxy for topic and, indeed, it is a good proxy in specific domains such as museum tours. However, it does mean that the task domain and especially the corpus of notes from which suggestions are drawn must be chosen carefully to insure physical context will be useful for automatic retrieval of useful information.

3 Conclusion

Jimminy is an example where physical context, namely, location and people in the vicinity, is not especially useful for automatically retrieving information from a personal notes archive. However, given that the experiment described was conducted with only one subject, it is dangerous to generalize these results too far. As was stated in the introduction, the specific circumstances in which notes are taken will have a major impact on their structure and, thus, the usefulness of any one feature. What can be generalized are the lessons for identifying where physical context might be useful in such a system.

First, features should neither be so sparse that there are a large number of unique occurrences of a feature nor should they be clumped. Sparseness is a problem because it is unlikely that a match will be found at all for a sparsely distributed feature. Clumping is a problem because the feature will not distinguish between potentially useful suggestions. The concept that rare but nonunique features are best for retrieving information is well understood in the information-retrieval field and is used in most search-engine algorithms today [20].

Second, physical features need to correlate well with the topic in which a person is currently interested and the topic of a note that might be suggested. For example, the person feature in a salesman's notes may correlate very strongly with the topic because sales activities tend to center on people. Location may correlate well with topic in the notes of a tourist, museum patron, or police officer.

Finally, designing just-in-time information retrieval agents requires a good understanding of both the contexts in which they will be used (that is, the "query" the system will use) and the corpus from which suggested information will be drawn. By understanding the task domain, the system can be based on features that correlate well with the underlying topic of a note or user's environment.

Acknowledgments

The author would like to thank Professor Thad Starner for his feedback and advice throughout this research.

References

[1] P.E. Hart and J. Graham, "Query-Free Information Retrieval," *Proc. Conf. Cooperative Information Systems,* pp. 36-46, 1994.

[2] T. Selker, "Coach: A Teaching Agent that Learns," *Comm. ACM,* vol. 37, no. 7, pp. 92-99, 1994.

[3] E. Horvitz, J. Breese, D. Heckerman, D. Hovel, and K. Rommelse, "The Lumiere Project: Bayesian User Modeling for Inferring the Goals and Needs of Software Users," 1998.

[4] H. Lieberman, "Letizia: An Agent that Assists Web Browsing," *Proc. Int'l Joint Conf. Artificial Intelligence,* pp. 924-929, 1995.

[5] J. Budzik and K. Hammond, "Watson: Anticipating and Contextualizing Information Needs," *Proc. 62nd Ann. Meeting Am. Soc. for Information Science,* pp. 727-740, 1999.

[6] B. Rhodes and P. Maes, "Just-in-Time Information Retrieval Agents," *IBM Systems J.,* vol. 39, no. 3, pp. 685-704, 2000.

[7] P. Maglio, R. Barret, C. Campbell, and T. Selker, "Suitor: An Attentive Information System," *Intelligent User Interfaces,* ACM Press, pp. 169-176, 2000.

[8] S. Feiner, B. MacIntyre, and T. Höllerer, "A Touring Machine: Prototyping 3D Mobile Augmented Reality Systems for Exploring the Urban Environment," *Proc. First Int'l Symp. Wearable Computers,* pp. 74-81, Oct. 1997.

[9] G. Abowd, C. Atkeson, J. Hong, S. Long, R. Kooper, and M. Pinkerton, "Cyberguide: A Mobile Context-Aware Tour Guide," *ACM Wireless Networks,* vol. 3, pp. 421-433, 1997.

[10] J. Pascoe, "Adding Generic Contextual Capabilities to Wearable Computers," *Proc. Second Int'l Symp. Wearable Computers,* pp. 92-99, Oct. 1998.

[11] F. Sparacino, "The Museum Wearable: Real-Time Sensor-Driven Understanding of Visitors' Interests for Personalized Visually-Augmented Museum Experiences," *Proc. Museums and the Web (MW 2002),* Apr. 2002.

[12] R. Oppermann, M. Specht, and I. Jaceniak, "Hippie: A Nomadic Information System," *Lecture Notes in Computer Science,* vol. 1707, pp. 330-333, 1999.

[13] Y. Sumi, T. Etani, S. Fels, N. Simonet, K. Kobayashi, and K. Mase, "C-Map: Building a Context-Aware Mobile Assistant for Exhibition Tours," *Community Computing and Support Systems,* 1998.

[14] B. Schiele, T. Jebara, and N. Oliver, "Sensory Augmented Computing: Wearing the Museum's Guide," *IEEE Micro,* 2001.

[15] J. Rekimoto, Y. Ayatsuka, and K. Hayashi, "Augment-Able Reality: Situated Communications through Physical and Digital Spaces," *Digest of Papers: Second Int'l Symp. Wearable Computers,* pp. 68-75, Oct. 1998.

[16] A. Pentland and T. Choudhury, "Face Recognition for Smart Environments," *Computer,* vol. 33, no. 2, pp. 50-55, Feb. 2000.

[17] T. Starner, "Wearable Computing and Context Awareness," PhD dissertation, MIT Media Laboratory, Cambridge, Mass., May 1999.

[18] B. Rhodes and T. Starner, "Remembrance Agent: A Continuously Running Automated Information Retrieval System," *Proc. Practical Applications of Intelligent Agents and Multi-Agent Technology (PAAM),* Apr. 1996.

[19] B. Rhodes, "The Wearable Remembrance Agent: A System for Augmented Memory," *Personal Technologies J.,* special issue on wearable computing, vol. 1, no. 4, pp. 218-224, 1997.

[20] D. Harman, *Information Retrieval: Data Structures and Algorithms,* pp. 363-392. Prentice Hall, 1992.

[21] T. Starner, D. Kirsch, and S. Assefa, "The Locust Swarm: An Environmentally-Powered, Networkless Location and Messaging System," Technical Report 431, MIT Media Lab, Perceptual Computing Group, Apr. 1997.

[22] N. Minar, M. Gray, O. Roup, and P.M. Raffi Krikorian, "Hive: Distributed Agents for Networking Things," ASA/MA '99, 1999, http://hive.media.mit.edu/.

[23] B. Rhodes, N. Minar, and J. Weaver, "Wearable Computing Meets Ubiquitous Computing: Reaping the Best of Both Worlds," *Proc. IEEE Int'l Symp. Wearable Computers,* pp. 141-149, 1999.

[24] W. Newman, M. Eldridge, and M. Lamming, "PEPYS: Generating Autobiographies by Automatic Tracking," *Proc. European Conf. Computer Supported Cooperative Work,* pp. 175-188, 1991.

[25] M. Eldridge, M. Lamming, and M. Flynn, "Does a Video Diary Help Recall?" *Proc. Conf. People and Computers VII,* A. Monk, D. Diaper, and M.D. Harrison, eds., pp. 257-269, 1992.